

# **Language, Culture, and Communication**

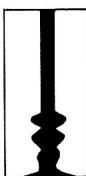
**Essays by Joseph H. Greenberg**

**Selected and Introduced  
by Anwar S. Dil**

**Stanford University Press, Stanford, California**

Language Science and National Development

A Series Sponsored by the  
Linguistic Research Group of Pakistan



General Editor: Anwar S. Dil

Stanford University Press  
Stanford, California

© 1971 by the Board of Trustees of the  
Leland Stanford Junior University

Printed in the United States of America

ISBN 0-8047-0781-2

Original edition 1971

Last figure below indicates year of this printing:

83 82 81 80 79 78 77 76 75 74

## **Contents**

Acknowledgments	vii
Introduction by Anwar S. Dil	xi
1. Linguistics and Ethnology (1948)	1
2. Historical Linguistics and Unwritten Languages (1953)	11
3. Concerning Inferences from Linguistic to Nonlinguistic Data (1954)	49
4. The Measurement of Linguistic Diversity (1956)	68
5. Language and Evolutionary Theory (1957)	78
6. Language, Diffusion and Migration (1957)	93
7. Language and Evolution (1959)	106
8. African Languages (1960)	126
9. African Tongues and Tribes (1960)	137
10. Memorandum Concerning Language Universals (1961)	143
11. Language and Linguistics (1963)	156

12.	Culture History (1964)	168
13.	The Study of Language Contact in Africa (1964)	185
14.	Urbanism, Migration and Language (1965)	198
15.	The History and Present Status of African Linguistic Studies (1965)	212
16.	Interdisciplinary Perspectives in African Linguistic Research (1966)	228
17.	Sociolinguistic Research in Africa (1966)	249
18.	The Nature and Definition of Language (1968)	260
19.	The Science of Linguistics (1968)	274
20.	Language Universals: A Research Frontier (1969)	295
21.	Linguistics as a Pilot Science (1971)	314
22.	Is Language like a Chess Game ? (1971)	330
	Bibliography of Joseph H. Greenberg's works Compiled by Anwar S. Dil and Emily Hallin	353

# 2

## Historical Linguistics and Unwritten Languages

### I. Historical Linguistics and Descriptive Linguistics

Unlike some other aspects of anthropology affected by the functionalist attack on history, the validity and fruitfulness of the historic approach in linguistics has never been seriously questioned. The objections which have been raised to certain assumptions of classical Indo-European comparative linguistics, such as the existence of sound laws without exceptions or the overliteral interpretation of the family-tree metaphor of language relationship, have not involved any fundamental doubt as to the legitimacy and value of historical reconstruction as such; at the most, they have, in the case of the Italian group of neo-linguists,<sup>1</sup> suggested specific alternative reconstructions of certain Proto-Indo-European forms.

The possibility of the application of traditional Indo-European methods to "primitive" (i. e., unwritten) languages has been deprecated by some Indo-Europeanists (Vendryes, 1925). It is evident that, while in principle the same procedures are appropriate, the absence of direct documentation for earlier historic periods is a distinct methodological handicap. The last decades, however, have seen the successful employment of classical reconstruction methods in a number of areas, including Central Algonkian by L. Bloomfield, Bantu by C. Meinhof, and Malayo-Polynesian by O. Dempwolff. It should be borne in mind that in all these cases we have rather closely related forms of speech, so that the task involved is more comparable to the reconstruction of Proto-Germanic or Proto-Slavic than that of Proto-Indo-European. These attempts do furnish an important demonstration of the universal scope of those mechanisms of linguistic change which were already known to

function in the more restricted area of the traditionally studied Indo-European, Finno-Ugric, and Semitic stocks (Hockett, 1948).

Much more serious than skepticism regarding the possibility of linguistic reconstruction in the absence of early written records is the widely held opinion, which will be discussed in a later section of this paper, that remote relationships or even those of the order existing within the Indo-European family cannot be established for primitive languages because of the far-reaching influence which one language can exercise on another even in fundamental traits of grammatical structure. It is even claimed that the genetic question here loses its meaning, in that one language can go back to several distinct origins and cannot therefore be said to belong to one family more than to another (Boas, 1920). It is worth observing that even in these cases the value of historic investigation is not denied as providing evidence of specific contacts, even though, it is held, the genetic question cannot be resolved. Thus Uhlenbeck, who, in his later writing, takes the view of genetic connections just mentioned, has lavished much time and effort on an attempt to show resemblances between the Uralic languages and Eskimo which require a historical explanation, while avoiding commitment as to the nature of the historic relationship involved.

While historic linguistics thus continues as a legitimate and major area of linguistic endeavor, it is undeniable that, with the rise of structural schools in European and American linguistics, the center of interest has shifted in the recent period from the historical problems which dominated linguistic science in the nineteenth century to those of synchronic description. The present preoccupation with descriptive formulations, which appears to be the linguistic analogue of the rise of functionalism, can contribute much that is valuable to diachronic studies. Most obviously, perhaps, any advance in descriptive techniques, by improving the quality of the data which constitute the basis of historical investigation, can furnish material for hypotheses of wider historical connections and likewise increase the precision of reconstruction for those already established. Another factor of great significance is the influence of the fundamental approach to language which all structur-

alists share, whatever their other divergences, namely, the concept of languages as a system of functional units. In its diachronic aspect this provides us with a view of change as related to a system and at least partially explainable in terms of its internal functioning through time. In the realm of sound patterns, some of these implications have been realized for some time. Thus Trubetskoy, as well as others, has distinguished between those sound changes which affect the sound structure of the language and those which leave it unchanged (Jakobson, 1931). This clearly parallels the synchronic distinction between phonetic and phonemic sound differences. Under the influence of this manner of thinking, sound change in language is more and more considered in terms of the shifts and realignment it produces in the sound structure of language rather than as a haphazard set of isolated changes, as in the traditional handbooks of historical linguistics.<sup>2</sup> The more rigorous formulation of alternations in the phonemic shape of morphemes (morphophonemics) has also borne fruit in Hoenigswald's exposition of the bearing of such data on internal reconstruction, that is, the reconstruction of certain aspects of the former states of a given language without resort to either related languages or historical records (Hoenigswald, 1950). Although historical linguists had in effect used this method without formulation, the emphasis on rigorous formulation of assumptions is, on the whole, beneficial in an area, such as historical reconstruction, in which it has so largely been lacking.

Although there is thus no fundamental opposition between the historical and descriptive approaches to language, the focusing of attention on synchronic problems in the recent historic period, combined with the traditional concentration of linguistic forces in the areas of a few major Eurasian speech families, has led to comparative neglect of the basic problems of historical research in unwritten languages.

## II. The Establishment of Linguistic Relationship

The fundamental achievement of nineteenth-century science in linguistics, as in certain other areas, notably biology, was to

replace the traditional static interpretation of similarities in terms of fortuitous coincidence among species as kinds, all of which were created at the same time and could vary only within fixed and narrow limits, with a dynamic historic interpretation of similarities as reflecting specific historical interrelationships of varying degrees of remoteness. Taxonomy, the science of classification, thus was no longer the attempt to find essential features connecting certain things more closely than others as part of a divine plan but rather based itself on the selection of those criteria which reflected actual historic relationships. In the language of biology, it was the search for homologies rather than mere analogies. In spite of the fruitfulness of the Indo-European hypothesis and the further successes of similar hypotheses in establishing the Finno-Ugric, Semitic, and other families, the assumptions on the bases of which these first victories of linguistics as a science were obtained were never clearly formulated, and the extension of these methods to other areas of the world has suffered from the beginning from a lack of clarity regarding the criteria of genetic relationship, resulting, in almost every major area, in a welter of conflicting classifications and even in widespread doubt as to the feasibility of any interpretation of linguistic similarities in terms of historical connections. Yet assumptions which have been the very foundation on which the edifice of modern linguistics has been reared and which have helped give it a rigorousness of method and precision of result which are admittedly superior to those dealing with any other phase of human cultural behavior should not be lightly abandoned unless, of course, the data actually demand it. In what follows, an attempt is made to formulate the principles in accordance with which similarities in language can be given a historical interpretation. It is hoped that this will furnish the guiding principles on the basis of which problems in the subsequent sections referring to specific areas can receive a reasonable solution.

The fundamental assumption concerning language on the basis of which historical interpretation of linguistic similarities becomes possible seems to have been first explicitly formulated by the great Swiss linguist, Ferdinand de Saussure, in his Cours de linguistique générale, although its relevance for historical

problems is not there stated. According to de Saussure, language is a system of signs having two aspects, the signifiant and the signifié, equivalent, in the terminology of Bloomfield and of American linguists, to "form" and "meaning," respectively. Moreover, the relationship between these two aspects of the linguistic sign is essentially arbitrary. Given any particular meaning, there is no inherent necessity for any particular set of sounds to designate it in preference to any other. Although first stated in this manner by de Saussure, this assumption actually underlies the nineteenth-century hypotheses of linguistic relationships and represents essentially the solution accepted by all modern linguists of the controversy descending from the Greeks concerning the naturalness versus the conventionality of language. Given the arbitrariness of the relationship between form and meaning, resemblances between two languages significantly greater than chance must receive a historical explanation, whether of common origin or of borrowing.

This statement regarding the arbitrariness of the sign does need some qualification, in that there is a slight tendency for certain sounds or sound combinations to be connected more frequently with certain meanings than might be expected on a purely chance basis. Conspicuous instances are the nursery words for "mother" and "father" and onomatopoeias for certain species of animals. This is generally recognized as only a slight derogation from the principle of the arbitrariness of the sign, since the sound can never be predicted from the meaning; and, since such instances are relatively a minor factor from the point of view of frequency of occurrence, they will add slightly to the percentage of resemblances to be expected beyond those merely the result of chance between any two unrelated languages; but they are not adequate for the explanation of wholesale resemblances between two particular languages, such as French or Italian. Moreover, the few resemblances which rest on this factor can be allowed for by assigning them less weight in judging instances of possible historical connections between languages. This factor making for specific resemblances between languages will hereafter be called, 'somewhat inappropriately,' "symbolism," in accordance with the terminology employed by psychologists.

Given any specific resemblance both in form and in meaning between two languages, there are four possible classes of explanations. Of these four, two—chance and symbolism—do not involve historic relationship, in contrast to the remaining pair—genetic relationship and borrowing. These four sources of similarity have parallels in nonlinguistic aspects of culture. Genetic relationship corresponds to internal evolution, borrowing to diffusion, chance to convergence through limited possibilities (as in art designs), and symbolism to convergence through similarity of function.

Up to this point resemblances in form between two languages unaccompanied by similarity of meaning and those of meaning not bound to similarity of form have not been considered. I believe that such resemblances must be resolutely excluded as irrelevant for the determination of genetic relationship. They practically always arise through convergence or borrowing. Form without function (e.g., the mere presence of tonal systems or vowel harmony in two languages) or function without form (e.g., the presence of gender morphemes in two languages expressed by different formal means) is often employed as relevant for the determination of relationship, sometimes as the sole criterion, as in Meinhof's definition of Hamitic, or in conjunction with other criteria. The preference for agreements involving meaning without accompanying sound resemblances is sometimes based on metaphysical preconceptions regarding the superiority of form over matter (Kroeber, 1913).

Resemblance in meaning only is frequently the result of convergence through limited possibilities. Important and universal aspects of human experience, such as the category of number or a system of classification based on sex or animation in the noun or one of tense or aspect in the verb, tend to appear independently in the most remote areas of the world and can never be employed as evidence for a historical connection. That the dual number occurs in Yana (California), ancient Greek, and Polynesian is obviously an instance of convergent development. Sometimes semantic similarity without similarity in the formal means of expression is present in contiguous languages of similar or diverse genetic connection. In these cases we have the linguistic analogue of Kroeber's concept of

"stimulus diffusion"—indeed, a remarkably clear-cut instance of this process. Languages spoken by people in constant culture contact forming a culture area tend to share many such semantic traits through the mechanism of diffusion. This process may be carried to the point where it is possible to translate almost literally from one language to another. However, since it is precisely the semantic aspect of language which tends to reflect changes in the cultural situation and since such semantic resemblances cover continuous geographical areas, these resemblances are clearly secondary, however far-reaching they may be in extent. Beyond the inherent probabilities, there is much empirical evidence in areas from which documented history exists. Those traits which various Balkan languages share in common and which are one of the marks of the Balkans as a cultural area are largely semantic, involving a difference in the phonemic content employed as the mode of expression. Thus Rumanian, Serbian, and Greek express the future by "to wish" followed by an infinitive, but in Rumanian we have (1st person sing.) voiu + V, in Serbian ću + V, and in Greek tha + V. These are all known to be historically relatively recent and not a result of the more remote Indo-European genetic connections which all of them share. Roughly similar arguments hold for resemblances of form without meaning. There are limited possibilities for phonemic systems. For example, such historically unconnected languages as Hausa in West Africa, classical Latin, and the Penutian Yokuts share a five-vowel system with two significant degrees of length (a, a<sup>‡</sup>, e, e<sup>‡</sup>, i, i<sup>‡</sup>, o, o<sup>‡</sup>, u, u<sup>‡</sup>). Some resemblances in form without function are the result of the influence of one language on another, e.g., the clicks of Zulu which have been borrowed from the Khoisan languages. Normally, when related languages have been separated for a fairly long period, we expect, and find, considerable differences both in their sound systems and in their semantic aspects resulting from differential drift and the diversity of the cultural circumstances under which their speakers have lived. Too great similarities in such matters are suspect.

Since, as has been seen, resemblances in form without meaning and meaning without form are normally explainable by hypotheses other than genetic relationship, their presence does not

indicate, nor their absence refute, it. Hence they may be left out of consideration as irrelevant for this particular problem.

The evidence relevant to the determination of genetic relationship then becomes the extent and nature of meaning-form resemblances in meaningful elements, normally the minimal element, the morpheme. Lexical resemblance between languages then refers to resemblances in root morphemes, and grammatical resemblances refer to derivational and inflectional morphemes. The two basic methodological problems become the exclusion of convergence and symbolism, on the basis of significantly more than chance resemblance leading to a hypothesis of some kind of historical connection, and among these the segregation of those cases in which borrowing is an adequate explanation of the more-than-chance resemblances from those instances in which this is inadequate and genetic relationship must be posited.

The first approach to the problem of more than chance resemblances is quantitative. We may ask how many resemblances may be expected between any two languages which are not genetically related and have not borrowed from each other or from a mutual source. Several approaches seem possible. One would involve the calculation for each of the two languages of the expected number of chance resemblances on the basis of its phonemic structure and allowed phonemic sequences arranged in terms of what may be called "resemblance classes," based on a resolution as to what phonemes are to be considered similar to others for the purposes of the comparison. To such a procedure there are several objections. It does not eliminate the factor of symbolism, and it does not take into account the relative frequencies of the phonemes in each language. If, for example, in comparing two particular languages, it were agreed that the labials would all be treated as resembling one another and the dentals likewise and if, in both languages, dentals were five times as frequent as labials, the possibility of chance resemblance would be much greater than if they were equal. This objection could, of course, be met in principle by a weighting in terms of frequency, but in actual practice it would be difficult to carry out.

A more desirable procedure would be the following. Let us suppose that we have a list of one thousand morphemes matched for meaning in the two languages. In language A the first morpheme is kan, "one." Instead of calculating the abstract probability of a form resembling kan sufficiently to be considered similar, let us actually compare kan in form with all the thousand items on the other list. Let us likewise compare the meaning "one" with all the meanings on the other list. The chance probability of the existence of a form resembling kan, "one," in both form and meaning in list B will then be the product of form resemblances and meaning resemblances divided by 1,000, the total number of items. We should then do this for each morpheme in list A and total the probabilities. As can be seen, this is a very tedious procedure. Moreover, it will not include resemblances due to symbolism.

A much more practical method, which takes into account both chance and symbolism, is simply to take a number of languages which are admittedly unrelated and ascertain the number of resemblances actually found. The difficulty here is that results will vary with the phonetic structure of the languages. A number of such counts indicates that approximately 4 per cent is the modal value, employing a very generous interpretation of what constitutes similarity. Where, however, the two languages are similar in the phonemic structure of their morphemes, the degree of resemblance can become significantly larger. For example, between Thai and Jur, a Nilotic language, which have very similar phonemic structures, it reaches 7 per cent. It can be safely asserted that a resemblance of 20 per cent in vocabulary always requires a historical explanation and that, unless similarity of phonetic structure leads to the expectation of a high degree of chance similarity, even 8 per cent is well beyond what can be expected without the intervention of historical factors. This factor of the similarity or difference of the phonemic structure of morphemes is so important that in doubtful cases a simplified version of the second test, that of matching lists, should probably be applied. We might compare a particular form in list B with all those in list A from the phonemic point of view only, allowing merely one meaning, that of its partner in list A, presumably the nearest semantic equivalent. We then compare with the expected

frequency of resemblances (which is, of course, smaller than by the first method) only those cases of resemblances on the list in which the two forms are matched as nearest semantic equivalents. Thus, if as our first matching pair we had A nem, B kan, "one," and later in the list A ken, B sa, "only," the resemblance between A ken, "only," and B kan, "one," would be disregarded as not occurring in a matching pair.

In actual fact, however, this test can probably be dispensed with, since the mere quantity of resemblances in the form and meaning of morphemes is not the decisive factor in more doubtful cases. There are additional considerations based on the weightings to be accorded to individual items and the further fact that isolated languages are seldom found. The bringing-in of closely related languages on each side introduces new factors of the highest importance, which should lead to a definite decision.

Other things being equal, the evidential value of a resemblance in form and meaning between elements in two languages is proportional to the length of the item. A comparison such as A, -k; B, -k, "in," is, from this point of view at least, less significant than such a resemblance as A, pegadu; B, fikato, "nose." More important is the following consideration. The unit of comparison is the morpheme with its variant allomorphs, if these exist. If the two languages agree in these variations, and particularly if the variants are rather different in phonemic content, we have not only the probability that such-and-such a sequence of phonemes will occur in a particular meaning but the additional factor that it will be accompanied by certain variations in certain combinations. Agreement in such arbitrary morphophonemic variations, particularly if suppletive, i. e., involving no phonemic resemblance between the variants, is of a totally different order of probability than the agreement in a nonvarying morpheme or one in which the languages do not exhibit the same variation. Even one instance of this is hardly possible without historical connection of some kind, and, since, moreover, it is hardly likely to be borrowed, it virtually guarantees genetic relationship. We may illustrate from English and German. The morpheme with the main alternant hæv, "have," in English resembles the German chief allomorph ha:b, "have,"

both in form and in meaning. In English, hæv alternates with hæ- before -z of the third person singular present (hæz, "has"). In German, correspondingly, ha:b has an alternant ha- in a similar environment, before -t, indicating third person singular present, to form ha-t, "has." Likewise, English gud, "good," has the alternant be- before -tər, "comparative" and -st, "superlative." Similarly, German gu:t, "good," has the alternant be- before -sər, "comparative," and -st, "superlative." The probability of all this being chance, particularly the latter, which is suppletive, is infinitesimal. Since it is precisely such arbitrary variations, "irregularities" in nontechnical language, which are subject to analogical pressure, they tend to be erased in one or the other language, even if some instances existed in the parent-languages. Where they exist, however, they are precious indications of a real historical connection.

More generally applicable are considerations arising from the fact that the comparison is only in rare instances between two isolated languages. The problem as to whether the resemblances between two languages are merely the result of chance plus symbolism can then be tested by a number of additional methods. Let us say that, as is frequently the case, one or more other languages or language groups resemble the two languages in question but in the same indecisive way, that is, that this third or fourth language is not conspicuously closer to one than to the other of the two languages with which we have been first concerned. The following fundamental probability consideration applies. The likelihood of finding a resemblance both in form and in meaning simultaneously in three languages is the square of its probability in two languages. In general, the original probability must be raised to the n - 1 power where a total of n languages is involved, just as the probability of throwing a 6 once on a die is  $1/6$ , but twice is  $(1/6)^2$  or  $1/36$ . Similarly, if each of three languages shows a resemblance of 8 per cent to the other, which might in extreme cases be the result of mere chance, the expectation of the three languages all agreeing in some instance of resemblance in form and meaning will be  $(8/100)^2$  or  $64/10,000$ . In 1,000 comparisons, agreement among all three languages should occur only 6.4 times, that is, it will occur in 0.0064, or less than 1 per cent, of the comparisons. Hence a

number of instances of such threefold agreements is highly significant. If four or more languages which are about equally distant from one another agree in a number of instances, a historical connection must be assumed, and if this agreement involves fundamental vocabulary or morphemes with a grammatical function, genetic explanation is the only tenable explanation.

This may be illustrated from the Afroasiatic (Hamito-Semitic) family of languages consisting of five languages or language groups—Egyptian, Berber, Semitic, Chad (Hausa and others), and Cushite. The forms involved are guaranteed as ancestral in each group by the requirement of earliest attestation, as in the requirement for Egyptian that it occur in the Pyramid Texts, our oldest document, or of appearance in at least two genetic subgroups (as in the case of Chad and Cushite), so that, in effect, we are comparing five languages. Allowing again the very high total of 8 per cent of chance resemblance between any two of the languages, the expected number of occurrences of morphemes similar in form and meaning in all five groups simultaneously becomes  $(8/100)^4$  or  $2,816/100,000,000$ . Assuming that about 1,000 forms are being compared from each language, this leads to the expectation of  $2,816/100,000$  of a morpheme. That is, if one compared a series of five unrelated languages at random, employing 1,000 words in each case, the operation would lead to a single successful case in approximately 35 such sets of comparisons. As a matter of fact, eleven morphemes are found in the case of Hamito-Semitic instead of the expected  $1/35$ . There is only an infinitesimal probability that this could be the result of pure chance. In this case, the morphemes involved include such examples as -t, fem. sing. and -ka, second person singular masculine possessive. Genetic relationship, of which there are many other indications, seems the only possible explanation here.

Languages should never be compared in isolation if closer relatives are at hand. For the tendency of those particular forms in a language which resemble another language or group of languages to reappear with considerable frequency in more closely related forms of speech is a valuable index of the existence of a

real historical connection. The statistical considerations involved may be illustrated once more from the Hamito-Semitic family. The question whether Hausa is indeed related to Egyptian, Semitic, Berber, and the Chad language has always been treated through isolated comparisons between Hausa and the other groups, while the existence of more than seventy languages of the Chad group which show a close and obvious relation to Hausa has been ignored.

A comparison of basic vocabulary between Hausa and Bedauye, a contemporary language of the Cushite branch of Hamito-Semitic, shows 10 per cent agreement in vocabulary. It is clear that Hausa will have lost certain Proto-Hamito-Semitic words retained by Bedauye, and vice versa. The percentage of retained vocabulary is expressed by a simple mathematical relation, the square root of the proportion of resemblances. The proportion of Hausa vocabulary which is of Proto-Hamito-Semitic origin should therefore be  $\sqrt{10/100}$  or approximately  $32/100$ . If we now take another Chad language belonging to a different subgroup than Hausa, namely, Musgu, the percentage of resemblance to Hausa is 20 per cent. Applying the same reasoning, the percentage of Hausa vocabulary retained from the time of separation from Musgu, that is, from the Proto-Chad period, is  $\sqrt{20/100}$ , or approximately  $45/100$ . If, then, we take forms found in Hausa which resemble Egyptian, Berber, Semitic, or Cushite and because of the existence of a true genetic relationship these forms actually derive from Proto-Hamito-Semitic, they must also be Proto-Chad. Since Hausa has lost its forms since the Proto-Chad period independently of Musgu, which belongs to another subbranch, a true Proto-Hamito-Semitic form in Hausa should reappear by chance in Musgu  $32/100 \div 45/100$  of the time, that is  $32/45$ . On the other hand, if Hausa is not related to the other Hamito-Semitic languages, the apparent resemblances to them are accidental, and these words should reappear in Musgu no more frequently than any other, that is, 20 per cent of the time,  $9/45$  rather than  $32/45$ . An actual count shows that, of 30 morphemes in Hausa which resemble those of branches other than Chad, 22 occur in Musgu. This is  $22/30$  or  $33/45$ , remarkably close to the expected  $32/45$ . On the other

hand, of 116 forms which show no resemblances to those of other Hamito-Semitic branches, only 14 occur in Musgu.

Beyond the frequency of resemblances and their distribution in other languages of the same group, the form which the resemblances take is likewise of importance. If the resemblances are actually the result of historical relationship, even cursory reconstruction should show greater resemblance in most cases between the reconstructed forms than between those of two isolated languages. If the resemblances are all convergences, on the whole, reconstruction should increase the difference of the forms. This can be done in a tentative manner as the comparison proceeds and without necessarily involving the full apparatus of formal historical reconstruction, which is often not feasible with poor material or where the relationship is fairly remote and no written records are available. If, for example, we compared present-day Hindustani and English, we would be struck by a number of resemblances in basic vocabulary, including numerals, but the hypothesis of chance convergence would certainly appear as a plausible alternative. Even without going beyond contemporary Germanic languages, on the one hand, and Indo-Iranian languages, on the other, reconstruction would show a strong tendency to convergence of forms as we went backward in time, suggesting a real historical connection. Thus English tuwp resembles Hindustani dā:t only slightly. On the Germanic side comparison with High German tsa:n already suggests a nasal consonant corresponding to the nasalization of the Hindustani vowel. Conjecture of a possible \*tamp or the like as a source of the English and German form is confirmed by the Dutch tand. On the other hand, comparison of Hindustani with other Aryan languages of India suggests that the Hindustani nasalized and long vowel results from a former short vowel and nasal consonant, as in Kashmiri and Sindhi dand. Reconstruction has thus brought the forms closer together.

Last, and very important, a degree of consistency in the sound correspondences is a strong indication of historical connection. Thus, reverting to the English-Hindustani comparison, the presence of t in English tuw, "two," ten, "ten," and tuwp

"tooth" corresponding to Hindustani d in dō, das, and dā:t, respectively, is a strong indication of real historical relationship.

Assuming that such a relationship has been established, there still remains the problem of whether the resemblances in question can be explained by borrowing. While in particular instances the question of borrowing may be doubtful, I believe it is always possible to tell whether or not a mass of resemblances between two languages is the result of borrowing. The most important consideration is the a priori expectation and historical documentation of the thesis that borrowing in culture words is far more frequent than in fundamental vocabulary and that derivational, inflectional, pronominal morphemes and alternating allomorphs are subject to borrowing least frequently of all.

The oft repeated maxim of the superiority of grammatical over vocabulary evidence for relationship owes what validity it has to this relative impermeability of derivational and inflectional morphemes to borrowing. On the other hand, such elements are shorter, hence more often subject to convergence, and usually few in number, so that in themselves they are sometimes insufficient to lead to a decision. Lexical items are, it is true, more subject to borrowing, but their greater phonemic body and number give them certain compensatory advantages. While it cannot be said, a priori, that any single item might not on occasion be borrowed, fundamental vocabulary seems to be proof against mass borrowing. Swadesh, in a recent discussion of the problem of borrowing versus genetic explanations, presents quantitative evidence for the relative impermeability of fundamental vocabulary in several instances where the history of the language is known (Swadesh, 1951).

The presence of fundamental vocabulary resemblances well beyond chance expectation, not accompanied by resemblances in cultural vocabulary, is thus a sure indication of genetic relationship. This is a frequent, indeed normal, situation where a relationship is of a fairly remote order. Pronoun, body parts, etc., will agree while terms like "pot," "ax," "maize," will disagree. The assumption of borrowing here runs contrary to common sense and documented historic facts. A people so strongly influenced by

another that they borrow terms like "I," "one," "head," "blood," will surely have borrowed cultural terms also. Where the mass of resemblances is the result of borrowing, a definite source will appear. The forms will be too similar in view of the historical remoteness of the assumed relationship. Moreover, if, as is usual, the donor language is not isolated, the fact that the resemblances all point to one particular language in the family, usually a geographically adjacent one, will also be diagnostic. Thus the Romance loan words in English are almost all close to French, in addition to hardly penetrating the basic vocabulary of English. If English were really a Romance language, it would show roughly equal similarities to all the Romance languages. The absence of sound correspondences is not a sufficient criterion, since, where loans are numerous, they often show such correspondence. However, the presence of a special set of correspondences will be an important aid in distinguishing loans in doubtful instances. Thus French loan words in English show regular correspondences, such as Fr. s = Eng. c or Fr. ã = Eng. an (šás:čæns; šát: čænt; še:z:čejr, etc.).

Genetic relationship among languages is, in logical terminology, transitive. By a "transitive" relation is meant a relation such that, if it holds between A and B and between A and C, it must also hold between B and C. If our criteria are correct and languages do have single lines of origin, we should never be led by their application to a situation in which A appears to be related both to B and to C, but B and C themselves cannot be shown to be related. If this were so, A would consist equally of two diverse components, that is, would be a mixed language of elements of B and C. This situation is sometimes said to exist, and even on a mass scale. Africa is perhaps most frequently mentioned in this connection. Thus Boas (1929) writes: "... a large number of mixed languages occur in Africa. His [Lepsius'] conclusions are largely corroborated by more recent investigation of the Sudanese languages."

Close investigation shows that, of the hundreds of languages in Africa (800 is the conventional estimate), there is only one language concerning which the problem of genetic affiliation could

conceivably lead to two disparate classifications, the Mbugu language of Tanganyika. Even here the answer is clear that, in spite of the borrowing of Bantu prefixes and a large amount of vocabulary, mostly nonfundamental, the language belongs to the Cushite branch of Hamito-Semitic. The pronouns, verb forms, and almost all the fundamental vocabulary are Cushitic. The conventional African classification based on purely formal criteria, such as tone, combined with purely semantic, such as gender, had no connection with historical reality, and the necessarily contradictory results which followed led to the assumption of widespread mixture. If, as was done, we define a Sudanese language as monosyllabic, tonal, and genderless, and a Hamitic language as polysyllabic, toneless, and having sex gender, a polysyllabic, tonal language with sex gender (like Masai) will have to be interpreted as the result of a mixture of Sudanic and Hamitic elements.

The last full-scale treatment of this subject is Meillet's, which was followed by the counterarguments of Schuchardt, Boas, and others and a discussion of these objections by Meillet (1914). The present discussion is in fundamental agreement with Meillet in asserting that the genetic question always has a meaning and is susceptible of an unambiguous answer. Meillet differentiates between concrete grammatical resemblances involving both form and meaning and those involving meaning only without form, but only in passing. Similarly, he mentions rather casually the fact that fundamental vocabulary is not commonly borrowed, but does not exploit this insight. The advantages gained by collateral comparison with additional closely related languages, and the statistical significance of coincidences in three or more languages are not considered. The result is an unnecessarily skeptical attitude toward the possibilities of establishing genetic classification where there are no early written documents or where the grammatical apparatus is slight or nonexistent (e.g., Southeast Asia).

The objections of Schuchardt and Boas are in large part taken into account in the present analysis by the distinction between resemblances based on form and meaning which result from contact with other linguistic systems and those involving form only or meaning

only. It would perhaps be desirable to distinguish these by the terms "borrowing" and "influence," respectively. Justice is then done to Boas' insistence that diffusion is prominently operative in linguistic as in other cultural phenomena, by setting no limit to influence, which in the case of Creole language reaches its peak, while maintaining, in accordance with all the available evidence, that there are definite bounds to borrowing, since it tends to cluster in nonfundamental vocabulary and makes only rare and sporadic inroads into basic vocabulary and inflectional and derivational morphemes. What is commonly said about the grammatical effects of one language on another refers almost entirely to influence, not borrowing, in the sense of the terms as employed here.

In other words, the effects of one language upon another are extremely widespread, fundamental, and important. What is maintained here is merely that the results are of a kind that can be distinguished from those caused by genetic relationship. Nor is it asserted that the genetic affiliation of a language is the sole important historic fact concerning it. The effects of borrowing and influence, being more recent chronologically and giving specific insights into the nature of the contacts involved, may frequently be of greater significance to the ethnologist and culture historian than the factor of more remote genetic affiliation.

These two types of historical connections between languages are carefully distinguished by Trubetskoy. A group of languages which have affected one another by influence and borrowing and form a group analogous to a culture area is termed a Sprachbund, while a group of genetically linked languages is termed a Sprachfamilie. They become genera of the larger species, Sprachgruppe, taking in all types of historical connections between languages (Trubetskoy, 1928).

The common habit of confusing these two situations by the use of the term "mixed language," as though a language were a mechanical aggregate of a number of components which enter into it the same way but merely in different proportions that English is, say, 48 per cent Germanic, 43 per cent French, 4 per cent Arabic,

and 0.03 per cent Aztec (because of "tomato," "metate," etc.) is a gross oversimplification and fails to distinguish the different origin and function of the Germanic as opposed to the Romance-Latin and other components in English.

From what has been said, it should be evident that the establishment of genetic relationships among languages is no mere *jeu d'esprit*. It is the indispensable preliminary to a determination of the causes of resemblances between languages by leaving borrowing as the only remaining source where more than chance resemblance does not lead to a hypothesis of relationship. Where such a relationship is present, it provides the basis for separation of autonomous from foreign elements through reconstruction of the ancestral language. Without such reconstruction, an understanding of the process of change in language undergoes a severe limitation to those few areas of the globe in which documented materials concerning the earlier forms of languages exist.

### III. Selected Regional Sketches

#### A. Africa

The attempt to reduce the number of language families in Africa at all costs, leading to overambitious syntheses combined with a disregard of concrete resemblances in form and meaning between elements of language in favor of typological criteria, such as the presence of tone, noun classes, sex gender, monosyllabic roots, etc., has characterized African linguistic classification from the earliest systematic attempts (Lepsius, F. Müller, etc.) onward.

The dominant classification in England and the United States has been a kind of synthesis, varying in details with different writers, based chiefly on the investigations of Westermann on the Sudanic languages and Meinhof on the Hamitic. Clear statements of the basis of this classification can be found in Werner (1915) and

in Tucker (1940), as well as elsewhere. According to this view, there are three great indigenous language families in Africa—Sudanic, Bantu, and Hamitic, with Semitic as a separate but late intrusion and Bushman as possibly related to Sudanic. A disputed point has been the status of Hottentot, which most assign to Hamitic with Meinhof but which some classify with Bushman to form a Khoisan family, while others leave it independent or at any rate unclassified. Each of the three main families has its basic characteristics. Thus Sudanic is monosyllabic, tonal, lacks stress, grammatical gender, and all inflection, and places the genitive before the possessed noun. Hamitic, at the opposite extreme, is defined as polysyllabic, possessing Ablaut variation, having grammatical gender and inflection, lacking tone, and placing the genitive after the noun. In addition, it possesses the characteristic of polarity, which can best be illustrated by an example. The Somali language uses the same formative for the singular of the masculine and the plural of the feminine, while another element marks simultaneously the singular of the feminine and the plural of the masculine. Meinhof often expressed the opinion that the Bantu languages, which are assigned characteristics almost midway between the Sudanic and Hamitic families, were the result of a mixture of the two or, as he once expressed it, "had a Hamitic father and Sudanic mother" (Meinhof, 1912).

It is admitted that few languages exhibit the traits of one of these families in full purity. Deviations from the ideal pattern are attributed to influences of one family on the other. It is held that such intimate fusions may result that the choice of the fundamental component can in certain cases be made only by an arbitrary decision. Such mixed groups of languages are the Semi-Bantu, formed from Sudanic and Bantu; Nilo-Hamitic, a fusion of Sudanic with Hamitic; and, in the view of many, Hottentot, with a Sudanic-like Bushman element and a Hamitic element.

It is clear that by applying such criteria, which have no reference to the concrete relations between the form and the meaning of specific linguistic signs, Chinese is a Sudanic language and Old French is Hamitic. The latter, indeed, possesses a very

striking bit of polarity in the use of -s to indicate the nominative singular and plural accusative of the noun as opposed to a zero suffix indicating the accusative singular and nominative plural (e.g., murs: mur = mur: murs). In addition, it possesses gender, Ablaut, and all the other stated characteristics of Hamitic speech. On the other hand, we are led to a crowning absurdity, in that forms of speech that are probably mutually intelligible can be classified as genetically distinct. Thus Meinhof, in classifying the languages of Kordofan, west of the Upper Nile, paid no attention to any other factor than the existence or absence of class prefixes in the noun. Three of these languages—Tegele, Tagoy, and Tumele—are similar, probably to the point of mutual intelligibility. Meinhof (1915-19) states: "A comparison of vocabulary shows that the numerals [sc. of Tegele] completely agree with those of Tumele. Moreover they are for the most part identical with the Tagoy numerals. Besides, a number of word stems and some verb forms of Tegele are identical with Tagoy and Tumele. But the grammatical structure of the noun indicates that Tegele is a Sudanic language because noun classification is absent while Tagoy and Tumele have clear noun classes. Apparently there has been a mixture of two diverse elements."

The other classification which has enjoyed currency is that of A. Drexel, adopted with a few modifications by Schmidt and by Kiekers in their respective volumes on the languages of the world. The Drexel classification embodies an attempt to demonstrate Sprachenkreise in Africa parallel to the Kulturkreise of the Graebner-Schmidt culture-historical school. This involves such violence to linguistic facts as the separation of the closely knit Mandingo group of languages into two unrelated families and the assumption of special Fulani-Malayo-Polynesian and Kanuri-Sumerian connections. There is no clear statement of the method employed in arriving at such conclusions.

The recent Greenberg (1949-50) classification concentrates on specific criteria which are relevant for actual historical relationship. The large heterogeneous Sudanic group, to which Westermann, in his more recent writings, denied genetic unity is split into a number of major and some minor stocks. The most important of those,

Westermann's West Sudanic, shows a genetic relationship to Bantu, as evidenced by a mass of vocabulary resemblances, agreement in noun-class affixes, and phonetic correspondences, including those relating to tone, to which Westermann himself had drawn attention and to which he had even attributed a genetic significance, without, however, modifying his general scheme of language families to take account of it. The Semi-Bantu languages show a special resemblance to the Bantu languages simply because they belong to the same subgroup of languages in the larger family, to which the name "Niger-Congo" is applied. Since these Semi-Bantu languages do not possess common features as against Bantu, the Bantu language must be classified as merely one of over twenty subgroups within that one of the fifteen branches of the vast Niger-Congo family which includes both Bantu and "Semi-Bantu" languages.

Other major independent families formerly classified as Sudanic are Central Saharan, Central Sudanic, and Eastern Sudanic. This latter family includes the so-called "Nilo-Hamitic" languages, along with the closely related Nilotc languages in a single subfamily.

Hottentot is treated along with the central Bushman languages as a single subgroup within the Khoisan languages, the other branches being Northern Bushman and Southern Bushman. The Khoisan languages, in turn, are related to Sandawe and Hatsa in East Africa to form a single Click family. Of Meinhof's various proposed extensions of Hamitic, Fulani is assigned to the westernmost subfamily of Niger-Congo; the "Nilo-Hamitic" languages (Masai, Nandi, etc.) are classed as Eastern Sudanic; and Hottentot belongs to the Click family. Hausa, along with numerous other languages of the Chad family, is put, along with the traditionally Hamitic Berber, Cushite, and Ancient Egyptian and with Semitic, into the Hamito-Semitic family, for which the name "Afroasiatic" is proposed, since there is no linguistic justification for granting Semitic a special status. The term "Hamitic," which has been the basis of much pseudo-historical and pseudo-physical reconstruction in Africa, is thus abandoned as not designating a valid linguistic entity. The Afroasiatic family thus consists of five co-ordinate

branches: (1) Berber, (2) Egyptian, (3) Semitic, (4) Cushite, and (5) Chad.

The Greenberg classification assumes a total of sixteen independent families in Africa. There is some possibility of a reduction in this total. The hypotheses of a Kunama-Eastern Sudanic and a Songhai-Niger-Congo relationship, in particular, are worth investigating.

Westermann has indicated his adherence to this new classification in all essentials and is expected to espouse it in a forthcoming article in the journal Africa.<sup>3</sup>

#### B. Oceania

There is general agreement on the existence of only two extensive groups of related languages in Oceania—the Malayo-Polynesian and the Australian. The remaining families are the Tasmanian and a whole series of unrelated language families in New Guinea and neighboring islands, to which the cover-name "Papuan" is applied, with the general understanding that there is no proof or even likelihood that these languages form a single stock. Regarding Malayo-Polynesian, there is general consensus concerning which languages are to be included in the family, and the historical work of reconstruction of the ancestral Malayo-Polynesian and other languages will be considered in the following section on "Southeast Asia."

For the other large group, the Australian languages, although the existence of widespread relationships within the continent is asserted by all investigators, there is lack of unanimity regarding the number of families, some maintaining the unity of Australian languages and others denying it.

The linguists of the period before W. Schmidt's important work were acquainted almost exclusively with the languages of the large group which covers all the south and much of the north of the

continent and ignored or were unaware of certain languages of the extreme northwestern and north-central parts of Australia which differ considerably from the great mass of Australian languages. These observers, therefore, assumed the unity of all Australian languages and were concerned chiefly with hypotheses of outside connections, with Africa, with India (Dravidian), or, in the case of Trombetti, with an Australian-Papuan-Andamanese group. This latter attempt, like all the others, proved abortive in this instance, if for no other reason than that the Papuan member is no linguistic unit of any sort (Ray, 1907).

It was Schmidt (1913, 1914, 1917-18) who laid the foundations of a more careful study of the problem in a series of articles in Anthropos, later republished as Die Gliederung der australischen Sprachen (1919). Schmidt distinguishes two main families of Australian languages: the southern, which covers approximately the southern two-thirds of the continent, and a northern. He explicitly denies the existence of a genetic relationship between these two groups. Unlike the southern family, which constitutes a true genetic unity, the northern, according to Schmidt, is not a family at all but consists of numerous diverse, unrelated forms of speech. In the light of clear statements to this effect, it is difficult to know what is meant in a historical sense by Schmidt's threefold division of these northern languages into those whose words end in consonants as well as vowels, those whose words end in vowels only, and those whose words end in vowels and liquids but not in other consonants. This last group occupies, according to Schmidt, an intermediate position between the other two, probably through a process of language mixture. This threefold division of the northern languages, as well as the separation into a northern and a southern family, seems strongly motivated by an attempt at correlation with the Kulturkreise established in this area by the ethnological school of which Schmidt is a leading exponent. Kroeger (1924), in a review of Schmidt's work, criticized this division on the ground of obvious fundamental vocabulary resemblances between the northern and southern languages. He followed this up with a study of the distribution of common vocabulary items, which showed a sublime disregard in their distribution for the

fundamental east-west dividing line which Schmidt had drawn across the Australian continent.

In a series of articles in Oceania (1939-40, 1941-43), Capell made substantial contributions to our knowledge of the languages of the northwestern and north-central parts of the continent and also revealed the surprising fact that many of these languages had noun-prefix classes resembling those of the Bantu languages in Africa in their general functioning but, one should hasten to add, without specific resemblances to them in form and meaning. Capell asserts the fundamental unity of all Australian languages. He divides them into suffixing languages, roughly equivalent to Schmidt's southern family, and prefixing languages, corresponding to Schmidt's northern division. The criterion employed is existence of verb suffixes or prefixes to form tenses and moods and to indicate pronominal reference. It is admitted that the northern languages are, to some extent, suffixing also. Within the northern group we have, again, a threefold division on principles different from those of Schmidt. Groups with multiple noun classes, two classes, and no classes are distinguished. Capell admits, in effect, that this is not a genetic analysis. It leads, as he himself points out, to an inevitable cul-de-sac similar to that of Meinhof in Africa, cited above. We are confronted with a pair of languages—Nungali and Djämindjung—which are almost identical except that Nungali has noun classes and Djämindjung has none. A similar pair is Maung and Iwaidja. Concerning these latter, Capell observes: "It is safe to say, however, that had Iwaidja multiple classification, it would hardly be more than a dialect of Maung" (Capell, 1939-40, p. 420).

The solution suggested here is a simple one, if one keeps in mind a primary canon of classification, one so obvious that it would hardly seem to need statement, yet is frequently disregarded in practice. Languages should be classified on linguistic evidence alone. Among the irrelevancies to be excluded is the extent of the area in which the language is found and the number of speakers. There is no reason to expect that families of genetically equal rank should necessarily occupy territories approximately equal in extent.

Germanic and Tokharian are coordinate branches of Indo-European, but a greater contrast in territory and population could hardly be imagined. Germanic covers substantial portions of four continents and numbers hundreds of millions of speakers. Tokharian has no speakers at all, since it is extinct.

The extent of fundamental vocabulary resemblance, including pronouns, among all languages in Australia and the specific similarities in the noun prefixes which connect many north Australian languages provide sufficient evidence of a single Australian family. This family has numerous subgroups, certainly at least forty, of which the large southern subgroup is just one which has spread over most of the continent (including the Murngin languages in north-east Arnhemland and the languages of the western Torres Straits Islands). The ancestral Australian language had noun classes, and the southern subgroup has, like some of the northern languages (the prefixing, classless language of Capell's classification), lost these classes. It still maintains a survival, however, in the distinction of a masculine and a feminine singular pronoun found in certain southern languages in which the afformatives employed resemble those of the masculine and feminine singular classes among the class languages.

### C. Southeast Asia

There are sharp differences of opinion regarding linguistic relationships in this area. The following are the outstanding problems: (1) the validity of Schmidt's hypothesis of an Austroasiatic family consisting of Mon-Khmer, Munda, and other languages; (2) the validity of Schmidt's Austric hypothesis connecting Austroasiatic in turn with Malayo-Polynesian; (3) the affiliations of Thai and Annamite, connected by some with Chinese in one subbranch of the Sino-Tibetan family, while others place Thai with Kadai and Indonesian (Benedict) and Annamite with Austroasiatic (Schmidt and others); (4) the linguistic position of the Man (Miao-Yao) and Min-Hsia dialects spoken by aboriginal populations in China.

Accepting certain earlier suggestions and adding some of his own, Schmidt (1906) has proposed that the following groups of

languages are related to one another in his Austroasiatic stock: (1) Mon-Khmer, (2) the Palaung-Wa languages of the middle Salween, (3) Semang-Sakai, (4) Khasi, (5) Nicobarese, (6) the Munda group, (7) Annamite-Muong, (8) the Cham group. If we except Cham, which most writers consider Malayo-Polynesian, a conclusion which can hardly be doubted, then all these languages share numerous resemblances in fundamental vocabulary, extending to pronouns. Moreover, excepting Annamite, which has shed all its morphological processes, there are certain important derivational morphemes whose rather uncommon formal nature (infixes), combined with their basic functions in the grammar, absolutely excludes chance and makes borrowing a completely improbable explanation. I do not see how such coincidences as an infixed -m in the Mon of Burma and the languages of the geographically remote Nicobar Islands, both with agentive meaning, to mention only one of a number of such instances, can be the result of anything but genetic relationship.

Maspero has sought to demonstrate a close connection between Annamite and Thai, which he considers to be Sino-Tibetan. This case rests chiefly on the irrelevant argument from form only — the monosyllabism and tonicity of Annamite, in which it resembles Thai and Chinese. The extensive lexical resemblances to Thai, which hardly touch basic vocabulary, must be looked upon as mostly borrowing with some convergence. On the other hand, the mass of fundamental vocabulary points clearly in the direction of the Austroasiatic languages, and I do not see how any hypothesis of borrowing can explain it. If borrowed, the source is not evident, since Annamite now resembles one, now another, of the Austroasiatic languages. It often shows an independent development from a hypothetical reconstruction which can hardly be the result of anything but internal development from the ancestral Austroasiatic form. Thus Annamite mōt, "one," makes sense as an independent contraction from \*moyat, found in this form only in the distant Mundari language of India. The language geographically nearest to Annamite Khmer has muy, presumably <moy> with loss of final -at. Santali, the chief Munda language, has mit <\*miyat<\*moyat. The absence of the modest morphological apparatus of other Austroasiatic languages in Annamite cannot be used as an argument

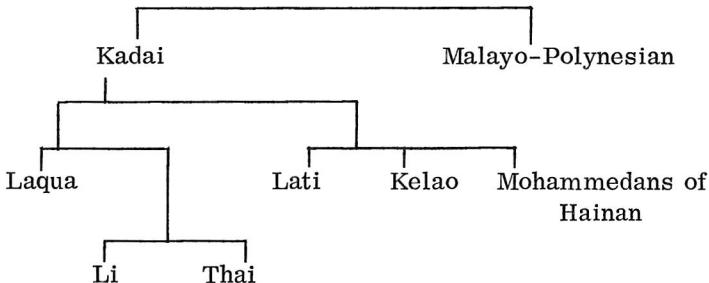
for any other relationship. The ancient maxim ex nihilo nihil fit may be appropriately applied in this instance.

Schmidt's further hypothesis of the relationship of Austroasiatic to the Malayo-Polynesian languages is of a far more doubtful nature. Most of the numerous etymologies proposed by Schmidt are either semantically or phonetically improbable or not attested from a sufficient variety of languages in one family or the other. Even with these eliminated, there remains a considerable number of plausible, or at least possible, etymologies, but very few of these are basic. Both language families employ prefixes and infixes, and the latter mechanism is certainly not very common. However, concrete resemblances in form and meaning of these elements which can reasonably be attributed to the parent-language of both groups are very few. Only pa-, causative, seems certain. In view of this, the Austric hypothesis cannot be accepted on present evidence. It needs to be reworked, using Dempwolff and Dyen's reconstructed Malayo-Polynesian forms, as well as taking into account the Thai and Kadai languages, which, as we shall see, are related to Malayo-Polynesian.

The traditional theory regarding Thai is that it forms, along with Chinese, the Sinitic branch of Sino-Tibetan. Benedict has proposed the relationship of Thai to the Kadai group, in which he includes certain languages of northern Indo-China, southern continental China, and the Li dialects of the island of Hainan. He has further posited the relationship of this Thai-Kadai family to Malayo-Polynesian (Benedict, 1942). Of the relation of Thai to the Kadai languages, which in the case of the Li dialects is particularly close, there can be no reasonable doubt. At the least, the traditional theory would have to be revised to include the Kadai languages, along with Thai, in Sinitic. I believe, however, that the connection of Thai with Chinese and Sino-Tibetan must be abandoned altogether and that Benedict's thesis is essentially correct. Thai resemblances to Chinese are clearly borrowings. They include the numerals from 3 on and a number of other words which are certainly the result of cultural contact. Thai is otherwise so aberrant that it must be at least another independent branch of Sino-Tibetan. Yet, when

resemblances are found, the forms are always like Chinese—altogether too like Chinese, one should add. Applying a test suggested earlier, it is found that those words in Thai which resemble Malayo-Polynesian tend to reappear in the Kadai languages, while those which are like Chinese do so only rarely. The proportion of fundamental vocabulary resemblances between Thai-Kadai and Malayo-Polynesian runs to quite a high number, far beyond chance and hardly explainable by borrowing, in view of the geographical distances involved.

I believe that Benedict's thesis needs restatement in some details of grouping, where, as so often happens, he has been led astray by nonlinguistic considerations, in this case the importance of Thai as a culture language. Thai shows special resemblance to the Li dialects of such far-reaching importance that Benedict's twofold division of Kadai into Laqua-Li and Lati-Kelao must be emended to put Thai along with Li in the first subgroup. In addition, the language of the Mohammedan population of Hainan does not belong, interestingly enough, with the Li dialects of the rest of the island but forms a third subdivision alongside the continental Lati-Kelao. The emended picture is shown in the accompanying diagram.



The Miao-Yao dialects of China have variously been called "Mon-Khmer" (i. e., Austroasiatic), "Sino-Tibetan," or "independent." There seems no good reason to classify them as other than a separate branch of Sino-Tibetan, no more divergent than, say, the Karen

languages of Burma. The evidence cannot be summarized here. The Min-Hsia language has been variously called a "Sino-Tibetan" or "Austroasiatic" language with a Chinese overlay. It likewise seems to be Sino-Tibetan. When the obvious Chinese borrowings are accounted for, the language still appears to show a special affinity to Chinese in fundamentals, so that it should probably be included in the Sinitic subbranch.

The question is here raised concerning the status of the Nehari language of India, classed by Grierson as Munda. It has been strongly influenced by Kurku, a neighboring Munda language; but, when allowance is made for this, the fundamental vocabulary and morphology of the language do not resemble those of any other family in the area. It may therefore be the only language of an independent stock. More material is needed to decide this question.

In summary, the language families of Southeast Asia are probably the following: (1) Sino-Tibetan, (2) Austroasiatic, (3) Kadai-Malayo-Polynesian, (4) Andaman Islands, (5) Nehari (?).

#### D. America North of Mexico

The present discussion is restricted to a few remarks of somewhat impressionistic character because of my lack of acquaintance with the linguistic data from this area. However, even cursory investigation of the celebrated "disputed" cases, such as Athabaskan-Tlingit-Haida and Algonkin-Wiyot-Yurok, indicate that these relationships are not very distant ones and, indeed, are evident on inspection. Even the much larger Macro-Penutian grouping seems well within the bounds of what can be accepted without more elaborate investigation and marshaling of supporting evidence. The difference between Oregon and California Penutian is comparable to that between any two of the subdivisions of the Eastern Sudanic family in Africa. The status of Algonkin-Mosan and Hokan-Siouan and the position of Zuni (which Sapir himself entered in the Azteco-Tanoan family with a query) strike me as the most doubtful points of Sapir's sixfold classification. The existence of a Gulf group, as

set forth recently by Haas, with a membership of Tunican, Natchez, Muskoghean and Timucua appears certain, as does the relationship of the Coahuiltecan languages both to the Gulf group and to the California Hokan in a single complex. Likewise, as Sapir pointed out, Yuki is probably no more than a somewhat divergent California Hokan language. The connection of Siouan-Yuchi and Iroquois-Caddoan with these languages is possible but far from immediately evident. Within Algonkin-Mosan, Salish-Chemakuan-Wakashan seems certain, as does Algonkin-Beothuk-Wiyot-Yurok (Beothuk may well be an Algonkin language). On the other hand, the relation of these two groups to each other and to Kutenai requires further investigation. Within the Azteco-Tanoan group it is clear that Kiowa is close to Tanoan and that Kiowa-Tanoan is related to Uto-Aztecán, as demonstrated by Trager and Whorf. The position of Zuni, as noted above, is very doubtful.

#### IV. Language and Historical Reconstruction

Ethnologists are rightly interested in comparative linguistic work, not so much for its own sake as for the light it sheds on other aspects of culture history. The basis for any discussion of this subject is inevitably the classic treatment of Sapir in his Time Perspective in Aboriginal American Culture. In spite of the brevity of this discussion, it is astonishingly complete, and there is little one would want to add to it, in spite of the lapse of time. The single most significant comment that might be made is that it serves as an essentially adequate basis for work in this field but that relatively little has been done toward the actual application of its principles. The problems involved are some of the most difficult in scientific co-operation and not easily solved. On the one hand, linguistic evidence is peculiarly suited to misapplication by ethnologists, who sometimes tend to use it mechanically and without at least an elementary understanding of the linguistic method involved. On the other hand, the linguist is often not greatly interested in problems of culture history, and the recent trend toward concentration in descriptive problems of linguistic structure draws him still further from the ordinary preoccupations

of archeologists and historically oriented ethnologists. Perhaps the ultimate solution is an intermediate science, ethnolinguistics, which will treat the very important interstitial problems, both synchronic and historical, which lie between the recognized fields of ethnology and linguistics.

The most important and promising recent development in this area is the possibility of establishing at least an approximate chronology for linguistic events in place of the relative time relations of classical historical linguistics. This method, known as "glotto-chronology" and developed chiefly by Swadesh and Lees, works on the assumption that rate of change in basic vocabulary is relatively constant. A chronological time scale is provided by comparisons of vocabulary from different time periods of the same languages in areas with recorded history. The results thus far indicate an average of ca. 81 per cent retention of basic vocabulary in one millennium. Thus, by comparing two related languages for which no earlier recorded material is available, the percentage of basic vocabulary differences will allow of an approximation of the date of separation of the two forms of speech.

By combining with this a rigorous application of Sapir's insight regarding the probable center of origin of a linguistic group, on the basis of a center of gravity calculated from the distribution of genetic subgroups, an instrument of historical reconstruction surpassing any previous use of linguistic data for these purposes becomes possible.

The center-of-gravity method may be briefly described as follows: Within each of the genetic subgroups of a linguistic family, the center of distribution is selected. If the subgroup is itself divided into clear dialect areas, the central point of each dialect area is calculated and the position of all is averaged to obtain the probable center of dispersal of the subgroup. The centers of the various subgroups are then averaged to obtain the most probable point of origin for the entire family. A correction in order to minimize the influence of single aberrant groups may be made by calculating a corrected center of gravity from the one reached by the above

method. The distance of the center of each subfamily is calculated from the center of gravity of the whole family. Then those subgroups which are most distant are weighted least, by multiplying the center of position of each subgroup by the reciprocal of the ratio of its distance to that of the most distant subgroup, and thus calculating a corrected value. Such results, mechanically arrived at, should, of course, be evaluated in terms of geographical and other collateral knowledge.

#### V. Goals, Methods, and Prospects

The goals and methods of comparative linguistics, particularly as applied to the field of primitive languages, are clear and generally agreed upon. The aims of this branch of science might be phrased in terms of the establishment of all possible genetic relationships between languages, the detection of all borrowings and the direction they have taken, and the maximal reconstruction of the ancestral languages which have given rise to the present languages. This is of value not only for its own sake and because these results can be employed toward general historical reconstruction but also because it gives us our basic knowledge of historic change in language under diverse circumstances. It is not until considerable data have been amassed in this field and a considerable variety of historical development in different areas has been traced that questions regarding overall change from one morphological or phonological type to another, leading to general laws of linguistic change, can ever be possible.

Problems of method, also, are in the main agreed upon. These resolve themselves into two main types: those pertaining to the determination of relationship and those concerning reconstruction. The latter problems are less controversial, and, in the United States at least, there is general agreement on the employment of what are essentially the procedures of classical Indo-European linguistics. The problems of establishing genetic relationships beyond the most self-evident ones, such as those of Powell in North America, admittedly involve more differences of opinion both

in Europe and in America. The abandonment of concrete criteria in favor of meaning without form or form without meaning and the abandonment of the traditional view regarding genetic relationship in some parts of the world in favor of the apparent profundity of analyses in terms of superposed strata have led only to increasing confusion and conflicting analyses, as they inevitably must. Moreover, only on the basis of clearly defined families established through specific form-meaning resemblances can reconstruction be attempted and with it the possibility of the study of historic process in language.

The greatest single obstacle to the rapid future growth of the field does not lie, however, in any conflict regarding aims or methods. It is rather the lack of trained people in sufficient number to provide the descriptive data for a vast number of languages, some of them near extinction. The topheavy concentration of linguistic scientists in the area of a very small number of language families of Eurasia and the extreme paucity of fully trained workers in such large areas as South America and Oceania are a grave handicap to future development of this field, as well as of linguistics as a whole. At the last meeting of the Linguistic Society of America, approximately 90 per cent of the papers presented on specific languages concerned a single language family, Indo-European.

The absence of effective liaison even between anthropological linguists and other branches of anthropology and its nonexistence in the case of other linguists, while an understandable consequence of the contemporary trend toward specialization, are likewise dangerous. Unless these situations are met and to some degree overcome, comparative linguistics must fall far short of the inherent possibilities afforded by the transparency of its material and the sophistication of its method of making a unique and significant contribution to the science of anthropology as a whole.

## NOTES

1. The reconstructions of the neo-linguistic school are not generally accepted by other scholars. For an exposition of neo-linguistic method, see G. Bonfante (1945). For a hostile critique see Robert Hall, Jr. (1946). It should perhaps be added that the approach of L. Hjelmslev in Denmark seems to exclude diachronic problems from language in principle but that this remains hardly more than a theoretic model.

2. Examples are the recent studies of Grimm's laws and other changes in Germanic by Twaddel and others, and various studies by Martinet of sound shifts (e.g., 1950).

3. Personal communication.

## REFERENCES

Benedict, P.

1942 "Thai, Kadai and Indonesian: A New Alignment in Southwestern Asia," American Anthropologist, XLIV, 576-601.

Boas, F.

1920 "The Classification of American Languages," American Anthropologist, XXII, 367-76.

1929 "Classification of American Indian Languages," Language, V, 1-7.

Bonfante, G.

1945 "On Reconstruction and Linguistic Method," Word, I, 83-94, 132-61.

- Capell, A.
- 1939-40 "The Classification of Languages in North and Northwest Australia," Oceania, X, 241-72, 404-33.
- 1941-43 "Language of Arnhem Land, North Australia," Oceania, XII, 364-92; XIII, 24-50.
- Greenberg, J. H.
- 1949-50 "Studies in African Linguistic Classification," Southwestern Journal of Anthropology, V, 79-100, 309-17; VI, 47-63, 143-60, 223-37, 388-98.
- Hall, Robert, Jr.
- 1946 "Bartoli's 'Neolinguistica,'" Language, XXII, 273-83.
- Hoenigswald, H.
- 1950 "The Principal Step in Comparative Grammar," Language, XXVI, 357-64.
- Hockett, C.
- 1948 "Implications of Bloomfield's Algonkin Studies," Language, XXIV, 117-31.
- Jakobson, R.
- 1931 "Principes de phonologie historique," TCLP, IV, 247-67.
- Kroeber, A. L.
- 1913 "The Determination of Linguistic Relationship," Anthropos, VIII, 389-401.
- 1924 "Relationship of the Australian Languages," Journal and Proceedings of the Royal Society of New South Wales, pp. 101-17.
- Martinet, A.
- 1950 "Some Problems of Italic Consonantism," Word, VI, 26-41.

- Meillet, A.
- 1914 "Le Problème de la parenté des langues,"  
Scientia, XV, No. XXXV, 3.
- Meinhof, Carl
- 1912 Die Sprachen der Hamiten. Hamburg: L. Friederichsen.
- 1915-19 "Sprachstudien im ägyptischen Sudan," Zeitschrift für Kolonialsprachen, VI, 161-205; VII, 36-133, 212-50, 326-35; VIII, 46-74, 110-39, 170-96, 257-67; IX, 43-64, 89-117, 167-204, 226-55.
- Ray, S.
- 1907 Linguistics. ("Reports of the Cambridge Anthropological Expedition to Torres Straits," Vol. III.) Cambridge: At the University Press.
- Schmidt, W.
- 1906 Die Mon-Khmer Völker. Braunschweig: F. Vieweg & Sohn.
- 1913 "Dei Gliederung der australischen Sprachen," Anthropos, VIII, 526-54.
- 1914 Ibid., IX, 930-1018.
- 1917 Ibid., XII, 437-39.
- 1918 Ibid., XIII, 747-817.
- 1919 Die Gliederung der australischen Sprachen. Vienna: Mechitharisten-Buchdruckerei.
- Swadesh, M.
- 1951 "Diffusional Cumulation and Archaic Residue as Historical Explanation," Southwestern Journal of Anthropology, VII, 1-21.
- Trubetskoy, N.
- 1928 Actes du premier Congrès International de Linguistes à La Haye. Leiden: Sijthoff.

- Tucker, A. N.  
1940      The Eastern Sudanic Languages. London: Oxford University Press.
- Vendryes, J.  
1925      Language: A Linguistic Introduction to History.  
Translated by Paul Radin. New York: A. Knopf.
- Werner, A.  
1915      The Language Families of Africa. London:  
Society for Promoting Christian Knowledge.